DESCRIPTION

INFORMATION PROCESSING APPARATUS AND METHOD, AND PROGRAM

5

TECHNICAL FIELD

The present invention relates to an information processing apparatus and method for executing a process on the basis of operation which is made based on input
10 speech with respect to an input form displayed on a display screen.

BACKGROUND ART

When data such as text or the like is input to
15 input fields (input forms) on a GUI (graphic user interface), an input form which is to receive an input is settled by selecting one of a plurality of input forms, and data is then input using a keyboard, or one of a plurality of candidates to be selected is selected
20 using a pointing device such as a mouse or the like to make an input. Also, upon inputting data to such input form, a technique for inputting data by means of speech using a speech recognition technique has been proposed.

However, according to the above prior art, upon
25 inputting data by speech to an input form, an input form which is to receive the data must be selected using a keyboard or mouse. Therefore, speech input and

manual input operations must be combined, and the operability is not always good.

DISCLOSURE OF INVENTION

5   The present invention has been made to solve the aforementioned problems, and has as its object to provide an information processing apparatus and method which can efficiently and flexibly execute operation to an input form displayed on a display window by input

10   speech, and a program.

In order to achieve the above object, an information processing apparatus according to the present invention comprises the following arrangement. That is, there is provided an information processing

15   apparatus for executing a process with respect to an input form displayed on a display screen on the basis of input speech, comprising:

storage means for storing input form information associated with the input form;

20   speech recognition means for recognizing the input speech; and

selection means for selecting input form information corresponding to a speech recognition result of the speech recognition means.

25   Preferably, the apparatus further comprises display control means for controlling a display pattern

of an input form corresponding to the input form information selected by the selection means.

Preferably, the input form information includes an input form name of the input form.

5   Preferably, the input form information includes layout information indicating a position of the input form.

Preferably, the display control means displays the input form corresponding to the input form

10   information selected by the selection means in a second display pattern which is different from a first display pattern of other input forms.

Preferably, the display control means displays the input form corresponding to the input form

15   information selected by the selection means at the center on the display screen.

Preferably, the apparatus further comprises informing means for, when selection by the selection means is settled, informing that message.

20   Preferably, the input form information includes an input form name of the input form, and layout information indicating a position of the input form,

the apparatus further comprises determination means for determining if the speech recognition result

25   of the speech recognition means corresponds to the input form name or the layout information, and

the selection means selects input form information corresponding to the speech recognition result of the speech recognition means on the basis of a determining result of the determination means.

5   Preferably, the input form information includes layout information indicating a position of the input form, and

the speech recognition means recognizes the input speech using speech recognition grammar data used to

10   recognize speech for specifying the layout information.

Preferably, the speech recognition grammar data includes data used to recognize at least one of a relative position expression indicating a relative position of the input form, and an absolute position

15   expression indicating an absolute position of the input form.

Preferably, the speech recognition grammar data includes data used to recognize if the absolute position expression corresponds to overall contents

20   including the input form or a display range on the display screen.

Preferably, when the input form is implemented by a hypertext document, the input form information includes a tag indicating the input form.

25   Preferably, the hypertext document describes a tag used to execute speech recognition by the speech recognition means.

In order to achieve the above object, an information processing method according to the present invention comprises the following arrangement. That is, there is provided an information processing method for

5   executing a process with respect to an input form displayed on a display screen on the basis of input speech, comprising:

a speech recognition step of recognizing the input speech; and

10   a selection step of selecting input form information associated with the input form, which corresponds to a speech recognition result of the speech recognition step.

In order to achieve the above object, a program

15   according to the present invention comprises the following arrangement. That is, there is provided an program for making a computer function as an information process for executing a process with respect to an input form displayed on a display screen

20   on the basis of input speech, comprising:

a program code of the speech recognition step of recognizing the input speech;

a program code of the selection step of selecting input form information associated with the input form,

25   which corresponds to a speech recognition result of the speech recognition step; and

a program code of the display control step of controlling a display pattern of an input form corresponding to the input form information selected in the selection step.

BRIEF DESCRIPTION OF DRAWINGS

Fig. 1 is a block diagram showing an example of the hardware arrangement of an information processing apparatus according to each embodiment of the present invention;

Fig. 2 is a functional block diagram of an information processing apparatus according to Embodiment 1 of the present invention;

Fig. 3 shows an example of an input form information table in Embodiment 1 of the present invention;

Fig. 4 shows the format of a recognition grammar in Embodiment 1 of the present invention;

Fig. 5 is a flow chart showing a process executed by the information processing apparatus of Embodiment 1 of the present invention;

Fig. 6 shows an example of a GUI in Embodiment 1 of the present invention;

Fig. 7 shows an example of a GUI in Embodiment 1 of the present invention;

Fig. 8 shows an example of a GUI in Embodiment 1 of the present invention;

Fig. 9 shows an example of a GUI in Embodiment 1 of the present invention;

Fig. 10 is a functional block diagram of an information processing apparatus according to

5 Embodiment 2 of the present invention;

Fig. 11 shows an example of an input form information table in Embodiment 2 of the present invention;

Fig. 12 is a flow chart showing a process

10 executed by the information processing apparatus of Embodiment 2 of the present invention;

Fig. 13 shows an example of a GUI in Embodiment 2 of the present invention;

Fig. 14 is a functional block diagram of an

15 information processing apparatus according to Embodiment 3 of the present invention;

Fig. 15 is a functional block diagram of an information processing apparatus according to Embodiment 5 of the present invention;

20 Fig. 16 is a flow chart showing a process executed by the information processing apparatus of Embodiment 5 of the present invention;

Fig. 17 shows an example of an input form information table according to Embodiment 6 of the

25 present invention; and

Fig. 18 shows an example of a tag used to execute speech recognition using a markup language according to Embodiment 7 of the present invention.

5 BEST MODE FOR CARRYING OUT THE INVENTION

Preferred embodiments of the present invention will now be described in detail in accordance with the accompanying drawings.

Fig. 1 is a block diagram showing an example of

10 the hardware arrangement of an information processing apparatus according to each embodiment of the present invention.

In the information processing apparatus, reference numeral 1 denotes a display device for

15 displaying a GUI. Reference numeral 2 denotes a central processing unit such as a CPU or the like for executing processes including numerical arithmetic operations·control, and the like. Reference numeral 3 denotes a storage device for storing temporal data and

20 a program required for processing sequences and processes of respective embodiments to be described later, or storing various data such as speech recognition grammar data, speech model, and the like. This storage device 3 comprises an external memory

25 device such as a disk device or the like, or an internal memory device such as a RAM·ROM, or the like.

Reference numeral 5 denotes a microphone for inputting speech uttered by the user. Reference numeral 4 denotes an A/D converter for converting speech data input via the microphone 5 from an analog signal into a digital signal. Reference numeral 6 denotes a communication device which exchanges data with an external device such as a Web server or the like via a network. Reference numeral 7 denotes a bus for interconnecting various building components of the information processing apparatus.

<Embodiment 1>

Fig. 2 is a functional block diagram of an information processing apparatus according to Embodiment 1 of the present invention.

Reference numeral 101 denotes a contents holding unit for holding contents to be displayed on a GUI, which is implemented by a hypertext document described using a description language (e.g., a markup language of an HTML document or the like). Reference numeral 102 denotes a GUI display unit such as a browser for displaying the contents held in the contents holding unit 101 on the GUI. Reference numeral 103 denotes a focus holding unit for holding an input form focused on various contents displayed on the GUI display unit 102. Reference numeral 104 denotes a form name generation unit for extracting input form names (notations) on the contents displayed on the GUI display unit 102, and

giving their pronunciations. The input form names and pronunciations generated by the form name generation unit 104 are held in a form name holding unit 105. In addition, the pronunciations are used as movement

5 recognition grammar data, and the input form names and pronunciations are held in a recognition grammar 106.

Fig. 3 shows an example of an input form information table which stores input form names (notations) and dictionary pronunciations in

10 correspondence with each other to manage information associated with input forms. In fig. 3, the dictionary pronunciations used for the input form information table is merely illustrative and other type of pronunciations can be used for the input form

15 information table.

Fig. 4 shows the format of the recognition grammar 106.

As shown in Fig. 4, the recognition grammar 106 comprises three kinds of speech recognition grammar

20 data including movement recognition grammar data used to select an input form to be focused by input speech, operation control recognition grammar data for various operations such as a response to confirmation to the user, a help request, and the like, and field value

25 recognition grammar data used to recognize contents input by speech to an input form. These speech

recognition grammar data may be combined into a single file or may form independent files.

Note that the speech recognition grammar data may include those which are normally used in speech
5   recognition, such as a word list that describes notations and pronunciations of words in case of single word speech recognition, a network grammar based on CFG (context-free grammar), and the like.

A description will revert to Fig. 2.

10   Reference numeral 107 denotes a speech input unit which comprises the microphone 5 and the A/D converter 4 for A/D-converting speech data input via the microphone 5. Reference numeral 108 denotes a speech recognition unit for reading out the speech recognition
15   grammar data held in the recognition grammar 106, and making speech recognition of a digital signal input from the speech input unit 107. Reference numeral 109 denotes a focus position change unit for, when the speech recognition result of the speech recognition
20   unit 108 indicates a given input form name, changing the focus position displayed on the GUI display unit 102 with reference to the focus holding unit 103.

The process to be executed by the information processing apparatus of Embodiment 1 will be described
25   below using Fig. 5.

Fig. 5 is a flow chart showing the process to be executed by the information processing apparatus of Embodiment 1 of the present invention.

Note that the operations of respective steps in the flow chart of Fig. 5 are stored, for example, as a program, in the storage device 3, and the central processing unit 2 reads out and executes that program.

In step S1, the GUI display unit 102 displays a GUI including a plurality of input forms to be displayed on the display device 1. The GUI may be displayed by loading and displaying external data such as HTML data which is described in a markup language, or may be displayed by only a dedicated program.

An example of the GUI will be described below using Fig. 6.

Fig. 6 shows an example of a GUI including a plurality of input forms to be displayed on the display device 1. This GUI assumes a registration (input/change) GUI of personal registration data as user information that pertains to a given user, and rectangular frames on Fig. 6 are respectively various input forms. For example, an input form 6 is used to input an ID number as character string data. Also, input forms 7, 9 to 13, and 15 to 22 are used to input various character string data. Input forms 8 and 14 are radio-button type input forms used to select desired choice data from those (male, female, business

man, and the like) prepared in advance. A button 23 is used to submit various data input to various input forms on the GUI to, e.g., an application.

When these input forms are displayed on the display device 1, the form name generation unit 103 generates their input form names and pronunciations, which are stored as an input form information table in the form name holding unit 104 and recognition grammar 106, as described above.

In case of a server-client type GUI display system including a Web server and a client that installs a Web browser, the process for generating the input form names, which is executed by the form name generation unit 103, may be executed in advance for respective contents on the Web server side, or may be dynamically executed on the Web browser on the client side.

In Embodiment 1, an input form which is to receive data (also referred to as a focused input form hereinafter) is indicated by the broken line (first display pattern), and a non-focused input form is indicated by the solid line (second display pattern). Fig. 6 exemplifies a case wherein the input form 6 is focused.

The registration GUI of personal registration data shown in Fig. 6 is an example for explaining a case wherein the personal registration data are to be

changed, and assume that personal registration data before change already exist. Upon changing the personal registration data, when the user inputs the ID number (e.g., 1234) to the input form 6 and presses the

5   submit button 23, as shown in Fig. 7, currently registered personal registration data corresponding to that ID number are displayed, and for example, the input form 9 is focused.

The description will revert to Fig. 5.

10   In step S2, the speech recognition unit 108 reads out various speech recognition grammar data from the recognition grammar 106 stored in the storage device 3. As described above, the speech recognition grammar data include the movement recognition grammar data used to

15   select an input form to be focused by input speech, operation control recognition grammar data, and field value recognition grammar data used to recognize speech input to the currently focused input form.

In step S3, the speech input unit 107 begins to

20   input speech. Speech uttered by the user is converted into an electrical signal by the microphone 5, and the electrical signal is further converted into a digital signal (speech data) by the A/D converter 4.

In step S4, the speech recognition unit 108

25   executes speech recognition of the input speech data using the read various speech recognition grammar data. In this case, speech recognition is made using the

movement recognition grammar data and field value
recognition grammar data, respectively. Since these
two speech recognition grammar data are used, speech
recognition results are obtained from the respective

5   speech recognition grammar data. These results are
compared using numerical values such as likelihood
levels that indicate the degrees of certainty of speech
recognition, and the speech recognition result with
higher degree of certainty is selected as a final

10   speech recognition result.

It is determined in step S5 if the speech
recognition result is selection of an input form. That
is, it is determined whether or not the likelihood of
the speech recognition result obtained using the

15   movement recognition grammar data is higher than that
of the speech recognition result obtained using the
field value recognition grammar data. If the speech
recognition result is not selection of an input form
(NO in step S5), the flow advances to step S8 to

20   display the speech recognition result of the speech
data input to the focused input form. Since this
process is the same as the prior art, a description
thereof will be omitted. On the other hand, if speech
recognition result is selection of an input form (YES

25   in step S5), the flow advances to step S6.

In step S6, an input form corresponding to the
speech recognition result (input form name) is selected.

For example, if an input form name "affiliation" or "address" is obtained as the speech recognition result, the flow advances from step S5 to this step S6, and an input form that matches the input form name

5 corresponding to the speech recognition result is specified. Fig. 9 shows an example of a GUI on the display device 1 when the speech recognition result is "affiliation".

In step S7, a selection confirmation operation is

10 made. This is a confirmation process for presenting the selected input form to the user. For example, display control for changing the display pattern of the selected input form to be distinguished from other non-selected input forms by flashing that input form

15 (changing the color of the form for a predetermined period of time) or the like is executed, or display control for scrolling a window to locate the selected input form at the center of the window or the like is executed. In addition, a beep tone may be produced to

20 indicate that the input form is selected.

As described above, according to Embodiment 1, when the user has uttered an input form name, an input form corresponding to the speech recognition result obtained by speech recognition of that utterance can be

25 selected as an input target of data. In this way, the user need not manually select an input form using a keyboard, mouse, or the like, and can select an input

form and input data with respect to the GUI by only input speech, thus improving the GUI operability compared to prior art.

<Embodiment 2>

5   In Embodiment 1, when the user has uttered an input form name, an input form as an input target of data is selected based on the speech recognition result obtained by speech recognition of that speech. Also, when the user has uttered a relative position

10   expression indicating a relative position of an input form, e.g., "third upper" or "second lower", an input form as an input target of data can be selected based on the speech recognition result obtained by speech recognition of that speech.

15   The functional arrangement of the information processing apparatus according to such embodiment is shown in Fig. 10.

       Fig. 10 is a functional block diagram of the information processing apparatus according to

20   Embodiment 2 of the present invention.

       Referring to Fig. 10, in addition to the contents holding unit 101, GUI display unit 102, recognition grammar 106, speech input unit 107, and speech recognition unit 108 in Fig. 2 of Embodiment 1, the

25   apparatus has a focus position change unit 109 for changing the focus position when the user has uttered the relative position expression, a focus position

holding unit 111 for holding the position of the
currently focused input form, a layout relationship
generation unit 112 for generating layout information
indicating input form names and their positions, a

5   layout relationship holding unit 113 for holding the
input form names and layout information held by the
layout relationship generation unit 112, and a relative
position determination unit 114 for determining if the
uttered contents are the relative position expression.

10   The input form names and layout information
generated by the layout relationship generation unit
112 are stored as an input form information table in
the storage device 3. Fig. 11 shows an example of that
table, which is managed as an input form information

15   table that stores the input form names and layout
information (e.g., vertical and horizontal position
coordinates when the upper left corner on the GUI is
defined as an origin) in correspondence with each other.
This input form information table is generated by

20   analyzing contents upon displaying the contents. When
contents are delivered from an external apparatus such
as a Web server or the like via a network, the input
form information table may be generated in advance on
the contents provider side, and may be submitted in

25   synchronism with submission of the contents. In
addition, in case of a server-client type GUI display
system including a Web server and a client that

installs a Web browser, the process for generating the
input form names and layout information, which is
executed by the layout relationship generation unit 112,
may be made in advance for respective contents on the
5 Web server side, or may be dynamically made on the Web
browser on the client side.

In Embodiment 2, the movement recognition grammar
data in the recognition grammar 106 in Fig. 11 contain
data required to make speech recognition of the
10 relative position expression, and data used to
recognize, for example, numerals, "th", "upper",
"lower", "right", "left", "from", and the like are
managed.

The process to be executed by the information
15 processing apparatus of Embodiment 2 will be explained
below using Fig. 12.

Fig. 12 is a flow chart showing the process to be
executed by the information processing apparatus of
Embodiment 2 of the present invention.

20 Note that Fig. 12 shows only different portions
from the flow chart of Fig. 5 of Embodiment 1.

When the speech recognition unit 108 executes
speech recognition of the input speech data with
reference to the read recognition grammar 106 in step
25 S4, the relative position determination unit 114
determines in step S70 if that speech recognition
result is a relative position expression. That is, it

is determined if the likelihood of the speech
recognition result obtained using the movement
recognition grammar data is higher than that of the
speech recognition result obtained using the field

5   value recognition grammar data. Especially, when the
likelihood of the speech recognition result obtained
using the movement recognition grammar data is higher
than that of the speech recognition result obtained
using other speech recognition grammar data, it is

10   determined that the speech recognition result is a
relative position expression.

     If it is determined in step S71 that the speech
recognition result is not a relative position
expression (NO in step S70), the flow advances to step

15   S8. On the other hand, if the speech recognition
result is a relative position expression (YES in step
S70), the flow advances to step S71, and the focus
position change unit 109 determines an input form
designated by the relative position expression. In

20   this case, the input form is determined using the
layout information of the currently focused input form,
the layout relationship holding unit 113, and the
speech recognition result of the relative position
expression.

25   For example, if the currently focused input form
is an input form 16 (Fig. 9), the focus position
holding unit 111 holds layout information (8, 1)

(Fig. 11) of the corresponding input form name "affiliation". If the speech recognition result of speech uttered by the user is "third upper", (5, 1) is determined as the movement destination of the focus

5 position on the basis of the input form information table in Fig. 11. In this way, the layout information held in the focus position holding unit 111 is updated to (5, 1). As a result, as shown in Fig. 13, the focus position is changed from the input form 16 to an input

10 form 12.

As described above, according to Embodiment 2, when the user has uttered a relative position expression that indicates the relative position of an input form, an input form corresponding to the speech

15 recognition result obtained by speech recognition of that utterance can be selected as an input target of data. In this manner, the user need not manually select an input form using a keyboard, mouse, or the like, and can select an input form and input data with

20 respect to the GUI by only input speech, thus improving the GUI operability compared to prior art. The user can select an input form by a simpler speech expression than Embodiment 1 without uttering an input form name, flexible, and precise input form selection by means of

25 input speech can be implemented.

<Embodiment 3>

In Embodiment 2, an input form is selected by the relative position expression. For example, an input form can also be selected by an absolute position expression indicating an absolute position such as

5   "fifth from top" or "second from bottom" uttered by the user.

The functional arrangement of the information processing apparatus according to such embodiment is shown in Fig. 14.

10   Fig. 14 is a functional block diagram of the information processing apparatus according to Embodiment 3 of the present invention.

Referring to Fig. 14, in addition to the contents holding unit 101, GUI display unit 102, recognition

15   grammar 106, speech input unit 107, and speech recognition unit 108 in Fig. 2 of Embodiment 1, and the focus position change unit 109, layout relationship generation unit 112, and layout relationship holding unit 113 in Fig. 10 of Embodiment 2, the apparatus

20   comprises an absolute position determination unit 121 and display range holding unit 122. The absolute position determination unit 121 implements a function similar to that of the relative position determination unit 114 in Fig. 10, and determines if the uttered

25   contents are an absolute position expression. Note that the details of the display range holding unit 122 will be explained later as Embodiment 4. The movement

recognition grammar data in the recognition grammar 106 contain data required to make speech recognition of the absolute position expression, and data used to recognize "from top", "from bottom", "from right",

5   "from left", numerals, "th", and the like are managed.

The process to be executed by the information processing apparatus of Embodiment 3 is an application of the process executed by the information processing apparatus of Embodiment 1. Especially, in the process

10   in step S6 of the flow chart in Fig. 5 of Embodiment 1, speech uttered by the user is recognized, and the absolute position determination unit 121 selects an input form to be focused with reference to the input form information table in Fig. 11. For example, when

15   the user has uttered "second from bottom", since the maximum value of the vertical position of the input form information table in Fig. 11 is 11, an input form of telephone number with the vertical position = 10 is selected, and the focus position is moved to that

20   position. After that, the flow advances to step S7.

As described above, according to Embodiment 3, an input form can be selected by the absolute position expression in place of the relative position expression, and more flexible, precise input form selection by

25   means of input speech can be implemented as in Embodiment 2.

<Embodiment 4>

When contents are browsed on a window application such as a browser or the like or on a portable device with a narrow display region, the GUI display unit 102 can only partially display contents, and the user must

5   scroll the contents to be browsed on the display window using a pointing device such as a mouse or the like. In each of the above embodiments, for example, when the user has uttered "third from top", the apparatus focuses on the third form from the top in the range of

10   the overall contents, but, for example, the apparatus may focus on the third form in the display range of the contents on the display window.

In such case, the display range holding unit 122 in Fig. 14 may hold layout information of the display

15   range currently displayed on the GUI display unit 102, and the absolute position determination unit 121 may determine the absolute position within the display range in the process in step S6 in Fig. 5.

When the user has explicitly uttered an absolute

20   position expression within the display range or that for the overall contents, either expression may be discriminated, and a corresponding operation may be made.   In this case, the movement recognition grammar data in the recognition grammar 106 in Fig. 14 contain

25   data required to make speech recognition of these absolute position expressions, and data used to manage, e.g., "overall", "within display range", and the like

are managed in addition to those described in Embodiment 3.

In this case, the absolute position of the overall contents or that within the display range in the display range holding unit 122 can be determined based on the speech recognition result if the user has designated like "third from top of overall" or "third from top in display range".

When designation indicating the absolute position of the overall contents or that within the display range is omitted, ambiguity occurs. In such case, either of these absolute positions may be fixed as a prescribed value, or the absolute position may be dynamically changed to the previously designated position. When the absolute position is dynamically changed, designation information for selecting either absolute position may be held in the display range holding unit 122.

The process to be executed by the information processing apparatus of Embodiment 4 is an application of the process executed by the information processing apparatus of Embodiment 1. Especially, in the process in step S6 of the flow chart in Fig. 5 of Embodiment 1, for example, if input forms 9 to 18 of the contents in Fig. 6 are displayed on the GUI display unit 102, the display range holding unit 122 holds an upper left

position (3, 1) and lower right position (9, 2) as layout information of the display range.

When the user has uttered "second from bottom", and especially when the display range holding unit 122 is set to adopt the display range as a default, the absolute position determination unit 121 determines layout information (8, 1) of input form name "affiliation" as the second input form from the bottom within the display range with reference to the display range holding unit 122 and the input form information table in Fig. 11, and moves the focus position to that position. On the other hand, when the display range holding unit 122 is set to adopt the overall contents as a default, the absolute position determination unit 121 determines layout information (10, 1) of input form name "telephone number" as the second input form from the bottom of the entire contents, and moves the focus position to that position.

As described above, according to Embodiment 4, the absolute position expression of the overall contents/absolute position expression within the display range is explicitly or automatically input by speech, and an input form can be selected by that input speech. In addition to the effects described in Embodiment 3, more flexible, precise input form selection by means of input speech according to the displayed display range can be implemented.

<Embodiment 5>

As the movement recognition grammar data in Embodiments 1 to 4 above, only data required to realize each embodiment are managed. Alternatively, the

5 movement recognition grammar data may be configured to be able to select an input form by any of the input form name, relative position expression, and absolute position expression.

The functional arrangement of the information

10 processing apparatus according to such embodiment is shown in Fig. 15.

Fig. 15 is a functional block diagram of the information processing apparatus according to Embodiment 5 of the present invention.

15 Referring to Fig. 15, in addition to the respective building components of Figs. 2, 10, and 14 of Embodiments 1 to 4, the apparatus has a position selection method determination unit 151 for determining the type (input form name, relative position expression,

20 and absolute position expression) of a focus position selection method.

The process to be executed by the information processing apparatus of Embodiment 5 will be described below using Fig. 16.

25 Fig. 16 is a flow chart showing the process to be executed by the information processing apparatus of Embodiment 5 of the present invention.

Note that Fig. 16 shows only portions different from the flow chart of Fig. 5 of Embodiment 1.

When the speech recognition unit 108 executes speech recognition of input speech data with reference to the read recognition grammar 106, the position selection method determination unit 151 determines with reference to the form name holding unit 105 in step S51 if the speech recognition result is selection of an input form. If the speech recognition result is selection of an input form (YES in step S51), the flow advances to step S61 to execute the same process as in step S6 in the flow chart of Fig. 5 of Embodiment 1. On the other hand, if the speech recognition result is not selection of an input form (NO in step S51), the flow advances to step S52.

The position selection method determination unit 151 determines in step S52 if the speech recognition result is a relative position expression. In this determination, for example, if the end of the speech recognition result is a position expression (e.g., "upper", "lower", "right", "left"), it is determined that the speech recognition result is a relative position expression.

If it is determined in step S52 that the speech recognition result is a relative position expression (YES in step S52), the flow advances to step S62 to execute the same processes as in steps S71 and S72 in

Fig. 12 of Embodiment 2. On the other hand, if the speech recognition result is not a relative position expression (NO in step S52), the flow advances to step S53.

5   The position selection method determination unit 151 determines in step S53 if the speech recognition result is an absolute position expression. In this determination, for example, if the head of the speech recognition result is a position expression (e.g.,

10   "from top", "from bottom", "from right", or "from left", or "of overall", "in display range", and their synonyms), it is determined that the speech recognition result is an absolute position expression.

If it is determined in step S53 that the speech

15   recognition result is an absolute position expression (YES in step S53), the flow advances to step S63 to execute a process for changing the focus position based on the absolute position expression, which has been explained in Embodiment 3 or 4. On the other hand, if

20   the speech recognition result is not an absolute position expression (NO in step S53), the flow advances to step S8.

In the description of Embodiment 5, the focus position can be selected by the selection method using

25   one of the input form name, relative position, and absolute position. Also, an arrangement capable of selecting the focus position using two or more

arbitrary selection methods of those methods can be implemented, needless to say.

As described above, according to Embodiment 5, in addition to the effects described in Embodiments 1 to 4,

5   since an input form can be selected by input speech by a plurality of types of selection methods, more flexible input form selection environment by means of input speech, which can be applied to various apparatuses, can be implemented.

10 <Embodiment 6>

When the contents held in the contents holding unit 101 are described using a markup language, the layout relationship holding unit 113 may hold the types of tags indicating input forms, and an input form may

15 be selected by input speech like "n-th (tag name)". Fig. 17 shows the contents of the input form information table held in the layout relationship holding unit 113 in such arrangement. In such case, the absolute position determination unit 121 recognizes

20 the first radio button as gender, and the second radio button as occupation. When the user inputs speech "second radio button", the focus position is moved to occupation, and the flow advances to step S7.

Note that the types of tags held in the layout

25 relationship holding unit 113 are not limited to "input" and "radio", and the same process can be made

if a "select" tag indicating a menu or an "a" tag"
indicating a link destination is held.

As described above, according to Embodiment 6,
since an input speech can be selected by input speech
5   in accordance with the type of tag indicating an input
form, more flexible input form selection by means of
input speech can be implemented.

<Embodiment 7>

When contents are described using a markup
10 language, there are many tags which are not used for
speech recognition inputs such as a "center" tag
indicating centering, "br" tag indicating a new line,
and the like.

Hence, in Embodiment 6, the types of tags used in
15 focus movement in speech recognition may be listed in a
portion that declares speech recognition.

Fig. 18 shows an example of tags used to execute
speech recognition using a markup language. In Fig. 18,
an example of tags of speech recognition associated
20 with Embodiment 7 is indicated, and the tags of speech
recognition [<SpeechRecog...>] is a description for
executing an input by speech recognition.

In the GUI display unit 102 in Embodiment 7,
[<SpeechRecog...>] is interpreted as "to make speech
25 recognition, and to display its speech recognition
result". The recognition grammar 106 used in speech
recognition, and a list of types of tags used in focus

movement in speech recognition can be designated by [grammar] and [used_tag], respectively. In this example, a tag [<SpeechRecog...>] declares that a recognition grammar dictionary [command.grm] is used,

5   and three different tags, i.e., "input" tag, "radio" tag, and "a" tag are used in focus movement.

As described above, according to Embodiment 7, since tags used to execute speech recognition are described in the contents together, the tags used to

10   execute speech recognition can be determined more efficiently in the tags in the contents. Also, since the tags used to execute speech recognition are described for each contents, even when an input form is selected by input speech in accordance with the type of

15   tag indicating an input form, the layout relationship holding unit 113 need not hold any input form information table in Fig. 17, and the storage resources can be saved.

Note that the present invention includes a case

20   wherein the invention is achieved by directly or remotely supplying a program (a program corresponding to the illustrated flow chart in each embodiment) of software that implements the functions of the aforementioned embodiments to a system or apparatus,

25   and reading out and executing the supplied program code by a computer of that system or apparatus. In such

case, the form is not limited to a program as long as the program function can be provided.

Therefore, the program code itself installed in a computer to implement the functional process of the present invention using computer implements the present invention. That is, the present invention includes the computer program itself for implementing the functional process of the present invention.

In this case, the form of program is not particularly limited, and an object code, a program to be executed by an interpreter, script data to be supplied to an OS, and the like may be used as along as they have the program function.

As a recording medium for supplying the program, for example, a floppy disk, hard disk, optical disk, magnetooptical disk, MO, CD-ROM, CD-R, CD-RW, magnetic tape, nonvolatile memory card, ROM, DVD (DVD-ROM, DVD-R), and the like may be used.

As another program supply method, connection may be established to a given home page on the Internet using a browser on a client computer, and the computer program itself of the present invention or a file, which is compressed and includes an automatic installation function, may be downloaded from that home page to a recording medium such as a hard disk or the like, thus supplying the program. Also, program codes that form the program of the present invention may be

broken up into a plurality of files, and these files may be downloaded from different home pages. That is, the present invention also includes a WWW server that makes a plurality of users download program files for

5 implementing the functional process of the present invention using a computers.

Also, a storage medium such as a CD-ROM or the like, which stores the encrypted program of the present invention, may be delivered to the user, the user who

10 has cleared a predetermined condition may be allowed to download key information that decrypts the program from a home page via the Internet, and the encrypted program may be executed using that key information to be installed on a computer, thus implementing the present

15 invention.

The functions of the aforementioned embodiments may be implemented not only by executing the readout program code by the computer but also by some or all of actual processing operations executed by an OS or the

20 like running on the computer on the basis of an instruction of that program.

Furthermore, the functions of the aforementioned embodiments may be implemented by some or all of actual processes executed by a CPU or the like arranged in a

25 function extension board or a function extension unit, which is inserted in or connected to the computer,

after the program read out from the recording medium is written in a memory of the extension board or unit.